

U.S. Appn. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

This listing of claims will replace all prior versions and listings of claims in the instant application:

### LISTING OF CLAIMS

1. (Currently Amended) In a text-to-speech system, a method of converting text-to-speech comprising:

receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

generating processed input by parsing and normalizing said text input;

~~a text-to-speech engine of the text-to-speech system processing said text input into processed input, said processed input comprising at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine;~~

comparing said processed input to at least one entry in a text-to-speech cache memory, wherein said entry in said text-to-speech cache memory specifies a corresponding spoken output, wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information, and wherein each spoken output has an assigned score; [[and]]

if said processed input matches one of said entries in said text-to-speech cache memory, providing said spoken output specified by said matching entry and rendering said spoken output according to said plurality of attributes associated with said text input;

if said processed input fails to match one of said entries, generating an additional spoken output with a text-to-speech engine, generating an entry that specifies said additional spoken output, assigning a score to said additional spoken output, storing said additional spoken output and assigned score in said cache memory, and rendering said

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

spoken output with the text-to-speech engine according to said plurality of attributes associated with said text input;

if the cache memory is full when said additional spoken output is generated, deleting from said cache memory a spoken output having a lower score; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of corresponding spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

2. (Previously Presented) The method of claim 1, wherein said text-to-speech cache entries include an intermediate output which is not a digitally encoded audio file; and wherein said text-to-speech engine converts said intermediate output to said spoken output.

3. (Previously Presented) The method of claim 1, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

4. (Original) The method of claim 1, further comprising:  
logging each said match of said text input with a text-to-speech cache entry.

5. (Currently Amended) The method of claim 1, wherein said text input does not match an entry in said text-to-speech cache memory, said method further comprising:  
determining a spoken output corresponding to said text input by using the text-to-speech engine to text-to-speech convert the text input[[ ; and ]]

(WP253724;1)

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

~~storing an entry in said text-to-speech cache memory corresponding to said text input, wherein said entry specifies said determined spoken output.~~

6. (Cancelled)

7. (Currently Amended) The method of claim 5, ~~wherein each said entry in said text-to-speech cache memory has a score, said method further comprising:~~  
~~periodically updating each said score.~~

8. (Cancelled)

9. (Cancelled)

10. (Currently Amended) The method of claim 9, ~~wherein said entries in said text-to-speech cache memory include attributes for customizing the spoken outputs said comparing step further comprising:~~  
~~comparing said attributes of said received text input with attributes of said entries in said text-to-speech cache memory.~~

11. (Currently Amended) A method of converting text-to-speech using a text-to-speech cache memory having a plurality of entries, wherein said entries comprise a processed form specifying a spoken output, wherein said processed form specifying spoken output does not comprise a digitally encoded audio file, said method comprising:  
receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

if said text input matches one of said entries in said text-to-speech cache memory, providing said processed form specified by said matching entry to a text-to-speech engine; [[ and ]]

said text-to-speech engine converting said processed form to said spoken output and rendering said spoken output according to said plurality of attributes associated with said text input; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

12. (Previously Presented) The method of claim 11, wherein the determined form of said text input comprises at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine.

13. (Previously Presented) The method of claim 11, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

14. (Original) The method of claim 11, further comprising:  
logging each said match of said text input with a text-to-speech cache entry.

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

15. (Cancelled)

16. (Cancelled)

17. (Cancelled)

18. (Cancelled)

19. (Currently Amended) A method of converting text-to-speech comprising:  
storing a plurality of entries in a text-to-speech cache memory, wherein the text-to-speech cache memory is directly and locally coupled to at least one text-to-speech engine, [[ and ]] wherein each said entry comprises a processed form specifying a spoken output, and wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information;  
assigning a score to each one of said plurality of entries;  
receiving a text input;  
processing said text input to determine a form specifying a spoken output for said received text;  
comparing said determined form of said text input with said entries in said text-to-speech cache memory;  
when at least one of the plurality of entries in said text-to-speech cache memory is matched to said determined form, retrieving the processed form for the matching entry from the text-to-speech cache memory, and using the processed form to generate said spoken output based on said attributes;  
when at least one of the plurality of entries in said text-to-speech cache memory is not matched to said determined form, using the at least one text-to-speech engine to generate said spoken output;

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

logging when one of said plurality of entries in said text-to-speech cache memory is matched to said received text input

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory; and

periodically updating said score for each one of said plurality of entries of said text-to-speech cache memory, wherein an updated score is computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score.

20. (Withdrawn) A method of administering entries of a cache memory comprising:

adding a plurality of entries to a cache memory and assigning a score to each one of said plurality of entries, wherein said scores are used to determine when a corresponding entry is deleted;

logging hits in said cache memory between a previous score update and a subsequent score update;

periodically updating each said score by multiplying each said score by a predetermined multiplier and adding a value representative of said logged hits for each one of said plurality of entries;

clearing said logged hits; and

deleting one of said plurality of entries in said cache memory having a lowest score.

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

21. (Currently Amended) A text-to-speech system comprising:  
a text-to-speech engine for receiving text [[ and ]] inputs and a plurality of attributes associated with said text and for producing a spoken output representative of said received text, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input; and  
a text-to-speech cache memory for storing selected entries corresponding to received text inputs, wherein said entries specify spoken outputs corresponding to said selected received text inputs, wherein at least one processing interaction occurs between the speech-to-text engine and the text-to-speech cache memory when the text-to-speech engine uses the text-to-speech memory cache to generate the spoken output responsive to receiving text, said processing interactions comprising at least one interaction selected from the group consisting of a pre-processing interaction where the received text is processed into an intermediate form before being compared to entries of the text-to-speech cache that are stored in said intermediate form and a post-matching interaction where the specified spoken outputs retrieved from the text-to-speech cache memory are processed by the text-to-speech engine to generate the spoken output according to the associated attributes.

22. (Cancelled)

23. (Previously Presented) The text-to-speech system of claim 21, wherein said text-to-speech cache entries include said spoken output, and wherein the processing interaction is a pre-processing interaction, and wherein the intermediate form comprises normalized text that represents a standardized version of the text input.

24. (Previously Presented) The text-to-speech system of claim 21, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

25. (Currently Amended) A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

generating processed input by parsing and normalizing said text input;

~~a text-to-speech engine of the text-to-speech system processing said text input into processed input, said processed input comprising at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine;~~

comparing said processed input to at least one entry in a text-to-speech cache memory, wherein said entry in said text-to-speech cache memory specifies a corresponding spoken output, wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information, and wherein each spoken output has an ordinal ranking; [[and]]

if said processed input matches one of said entries in said text-to-speech cache memory, providing said spoken output specified by said matching entry and rendering said spoken output according to said plurality of attributes associated with said text input;

if said processed input fails to match one of said entries, generating an additional spoken output with a text-to-speech engine, generating an entry that specifies said additional spoken output, assigning an ordinal ranking to said additional spoken output, storing said additional spoken output and assigned ordinal ranking in said cache memory,

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

and rendering said spoken output with the text-to-speech engine according to said plurality of attributes associated with said text input;

if the cache memory is full when said additional spoken output is generated, deleting from said cache memory a spoken output having a lower ordinal ranking; and

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of corresponding spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

26. (Previously Presented) The machine-readable storage of claim 25, wherein said text-to-speech cache entries include an intermediate output which is not a digitally encoded audio file; and wherein said text-to-speech engine converts said intermediate output to said spoken output.

27. (Previously Presented) The machine-readable storage of claim 25, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

28. (Original) The machine-readable storage of claim 25, further comprising:  
logging each said match of said text input with a text-to-speech cache entry.

29. (Previously Presented) The machine-readable storage of claim 25, wherein said text input does not match an entry in said text-to-speech cache memory, said method further comprising:

determining a spoken output corresponding to said text input by using the text-to-speech engine to text-to-speech convert the text input; and

(WP253724;1)

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

storing an entry in said text-to-speech cache memory corresponding to said text input, wherein said entry specifies said determined spoken output.

30. (Original) The machine-readable storage of claim 29, further comprising:  
removing one of said entries in said text-to-speech cache memory.
31. (Original) The machine-readable storage of claim 29, wherein each said entry in said text-to-speech cache memory has a score, said machine-readable storage further comprising:  
periodically updating each said score.
32. (Cancelled)
33. (Cancelled)
34. (Cancelled)
35. (Currently Amended) A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:  
storing a plurality of entries in a text-to-speech cache memory, wherein each one of said entries comprises a processed form specifying a spoken output wherein said processed form specifying spoken output does not comprise a digitally encoded audio file;  
receiving a text input and a plurality of attributes associated with said text input, wherein said attributes specify stress, gender, grammar, speed, and volume for an audio rendering of said text input;

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

if said text input matches one of said entries in said text-to-speech cache memory, providing said processed form specified by said matching entry to a text-to-speech engine; [[ and ]]

said text-to-speech engine converting said processed form to said spoken output and rendering said spoken output according to said plurality of attributes associated with said text input; and

generating a display of said text input wherein each word of said display is successsively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory.

36. (Previously Presented) The machine-readable storage of claim 35, wherein the determined form of said text input comprises at least one of normalized text that represents a standardized version of the text input and an intermediate format used by the text-to-speech engine.

37. (Previously Presented) The machine-readable storage of claim 35, wherein said text-to-speech cache is shared across multiple text-to-speech processes, wherein said text-to-speech processes are performed by a plurality of different text-to-speech engines, each engine utilizing said text-to-speech cache.

38. (Original) The machine-readable storage of claim 35, further comprising:  
logging each said match of said text input with a text-to-speech cache entry.

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

39. (Previously Presented) The machine-readable storage of claim 35, wherein said text input does not match an entry in said text-to-speech cache memory, said method further comprising:

determining a spoken output corresponding to said text input by using the text-to-speech engine to text-to-speech convert the text input; and

storing an entry in said text-to-speech cache memory corresponding to said text input, wherein said entry specifies said determined spoken output.

40. (Original) The machine-readable storage of claim 35, further comprising:  
removing one of said entries in said text-to-speech cache memory.

41. (Original) The machine-readable storage of claim 35, wherein each said entry in said text-to-speech cache memory has a score, said machine-readable storage further comprising:

periodically updating each said score.

42. (Original) The machine-readable storage of claim 41, further comprising:  
removing one of said entries in said text-to-speech cache memory having a lowest score.

43. (Currently Amended) A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

storing a plurality of entries in a text-to-speech cache memory, wherein the text-to-speech cache memory is directly and locally coupled to at least one text-to-speech engine, [[and]] wherein each said entry comprises a processed form specifying a spoken

(WP253724;1)

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

output, and wherein said text-to-speech cache memory contains a plurality of entries that specify spoken outputs, attributes for rendering spoken output, and callback information;

assigning a score to each one of said plurality of entries;

receiving a text input;

processing said text input to determine a form specifying a spoken output for said received text;

comparing said determined form of said text input with said entries in said text-to-speech cache memory;

when at least one of the plurality of entries in said text-to-speech cache memory is matched to said determined form, retrieving the processed form for the matching entry from the text-to-speech cache memory, and using the processed form to generate said spoken output based on said attributes;

when at least one of the plurality of entries in said text-to-speech cache memory is not matched to said determined form, using the at least one text-to-speech engine to generate said spoken output;

logging when one of said plurality of entries in said text-to-speech cache memory is matched to said received text input

generating a display of said text input wherein each word of said display is successively highlighted in coordination with an audible rendering of a word of said spoken output, coordination of said display and spoken output being based on call information stored in said cache memory; and

periodically updating said score for each one of said plurality of entries of said text-to-speech cache memory, wherein an updated score is computed by multiplying a previous score times a constant between zero and one and adding a number equal to the number of times a corresponding entry has been accessed since a last updating of the score.

{WP253724;1}

U.S. Appln. No. 09/941,301  
Amendment Dated Aug. 22, 2005  
Reply to Final Office Action of June 21, 2005  
Docket No. BOC9-2001-0022 (266)

44. (Withdrawn) A machine-readable storage, having stored thereon a computer program having a plurality of code sections executable by a machine for causing the machine to perform the steps of:

adding a plurality of entries to a cache memory and assigning a score to each one of said plurality of entries, wherein each said score determines when a corresponding entry is deleted;

logging hits in said cache memory between a previous score update and a subsequent score update;

periodically updating each said score by multiplying each said score by a predetermined multiplier and adding a value representative of said logged hits for each one of said plurality of entries;

clearing said logged hits; and

deleting one of said plurality of entries in said cache memory having a lowest score.

{WP253724;1}